

Rater Behaviour When Judging Language Learners' Pragmatic Appropriateness in Extended Discourse

Tetyana Sydorenko, Carson Maynard, Erin Guntly

The criteria by which raters judge pragmatic appropriateness of language learners' speech acts are underexamined, especially when raters evaluate extended discourse. To shed more light on this process, the present study investigated what factors are salient to raters when scoring pragmatic appropriateness of extended request sequences, and which specific aspects of performance they attend to as appropriate or inappropriate. Three judges evaluated request sequences using a 6-point scale, marked appropriate and inappropriate elements of each request, and explained how they approached the rating of each response. It was found that all raters oriented to the appropriateness of a request sequence as a whole, paying attention not only to the request proper but also to all follow-up moves, including appreciation and closing. Additionally, raters oriented to the surrounding context: the same expressions, such as a specific appreciation statement, were rated as appropriate in some contexts and inappropriate in others. Raters also oriented to pragmatic competence broadly, paying attention not only to appropriate pragmatic strategies and expressions in a particular context, but also to such aspects as intonation and cultural knowledge. Finally, while native and near-native speaker tendencies were observed, target speaker norms were not. Implications for pragmatics teaching and assessment are discussed.

Les critères selon lesquels les évaluateurs jugent la pertinence pragmatique des actes de langage d'apprenants de langue n'ont pas suffisamment fait l'objet d'études, notamment lors de l'évaluation de longues conversations. Pour éclairer davantage le processus, la présente étude a cherché à déterminer quels facteurs les évaluateurs jugent importants dans la pertinence pragmatique de séquences de requête étendues, et quels aspects spécifiques de la performance ils estiment appropriés ou pas. Trois juges ont évalué des séquences de requête selon une échelle de 6 points, ont indiqué les éléments appropriés et inappropriés de chaque requête et ont expliqué comment ils avaient abordé l'évaluation de chaque réponse. Les résultats indiquent que tous les évaluateurs jugeaient de la pertinence d'une séquence de requête dans son intégralité, portant attention non seulement à la requête comme telle mais aussi à toutes les démarches qui la suivaient, y compris le remerciement et la clôture. De plus, les évaluateurs tenaient compte du contexte : ils jugeaient qu'une même expression, une déclaration spécifique d'appréciation par exemple, était appropriée dans un contexte donné alors qu'elle ne l'était pas dans un autre. Ils ont également considéré la compétence pragmatique globale, notant, au delà des stratégies et des expressions pragmatiques appropriées dans un

contexte donné, des aspects comme l'intonation et les connaissances culturelles. Finalement, si les évaluateurs ont observé des tendances de locuteurs natifs ou quasi-natifs, on ne peut en dire autant des normes de la langue cible. On discute des incidences de l'étude sur l'enseignement et l'évaluation des compétences pragmatiques.

Assessment of second language (SL) pragmatics is understudied. The challenges of determining the construct representation, developing practical, valid, and reliable tests (e.g., Roever, 2008, 2011; Roever, Fraser, & Elder, 2014; Walters, 2007, 2009), and performing classroom assessment (Ishihara, 2013; Ishihara & Cohen, 2010) have been recently addressed in the literature, but one aspect that has received little attention is rater behaviour (for rare exceptions, see Alemi & Tajeddin, 2013; Brown & Ahn, 2011; Liu, 2007; Liu & Xie, 2014; Roever, 2008; Taguchi, 2006, 2011; Tajeddin & Alemi, 2014; Walters, 2007; Youn, 2007).

Additionally, studies that have examined rater behaviour generally elicited limited stretches of discourse; however, Roever (2008, 2011), Kasper (2006), and Yates (2010) stress the need to examine extended discourse. The present study adds to the limited knowledge in this area by examining extended discourse using computerized oral DCTs (discourse completion tasks). Finally, because rater judgments have been found problematic (e.g., Taguchi, 2006, 2011), some studies instead opted to compare pragmatic strategies used by learners and native speakers (NSs) (e.g., Otcu & Zeyrek, 2008; Rose, 2000; Taguchi, 2012). To further examine the benefits and challenges of the two approaches, we compared rater judgments with the categorization of discourse by pragmatic strategies (following Taguchi, 2006).

Studies of Rater Behaviour

Unreliability of rater judgments is a prominent issue in performance-based assessment (e.g., Brown, Hudson, Norris, & Bonk, 2002). Due to variation in pragmatic norms between NSs (Félix-Brasdefer & Koike, 2012), rating of pragmatics is even more subjective than that of other areas (Ishihara, 2013), and thus it is especially worth investigating.

Studies examining how raters approach the judgment of pragmatic performance cited such effects as different degrees of severity between raters and item types (Youn, 2007), differences in ratings between NS and non-native-speaking (NNS) judges (Alami & Tajeddin, 2013; Liu, 2007; Walters, 2007), familiarity of the raters with the learners' culture, and differential assignment of weightings to certain aspects of performance, such as grammar (Liu & Xie, 2014).

Several studies investigated whether rater training can counter the effects of raters' individual differences. Tajeddin and Alemi (2014) found that,

although training can increase interrater reliability as well as consistency in ratings within each judge, rater bias remained. Some raters continued to be more lenient, and others more severe, suggesting that training cannot completely remove the influence of raters' personal beliefs and experiences. Similar results were obtained by Brown and Ahn (2011). In contrast, Roevers (2008) found that despite a very short one-hour training period, the three raters were similar in their judgments. He attributed this finding to the simplicity of ratings: raters only needed to evaluate whether the response fit the situation rather than its overall appropriateness.

Although rater training does appear to increase agreement between raters, is it ecologically valid? As Taguchi (2011) and Wolfe and Chiu (1997) observed, even when given specific guidelines, some raters continued to rely on their own criteria. If one of the purposes of developing SL pragmatic competence is to appear pragmatically appropriate to NSs, perhaps the criteria of what is pragmatically appropriate should come from potential interlocutors with whom learners may interact, rather than from the researchers whose rating criteria are grounded in theories and, arguably, NS norms. To tackle this issue, Walters (2007) examined what raters actually pay attention to, when they are not given very explicit guidelines. Walters found that the NS rater sometimes relied on the knowledge of NS patterns, while the NNS rater sometimes took into account the examinees' fluency and clear pronunciation. Taguchi (2011) conducted a similar investigation into rater differences and found variability even after a norming session: raters differentially focused on linguistic forms, positive/negative politeness strategies, content, clarity of the message, and verbosity. Raters also weighed the criteria differently. Alemi and Tajeddin (2013) investigated the different rating criteria that NS and NNS ESL teachers apply, and found that NSs used 11 criteria, while NNS used only 6. NSs tended to use more specific criteria, such as "reasoning/explanation," while NNSs most commonly used the general "appropriateness" criterion.

Evaluating Speech Act Production via Strategies Used

Given that ratings of pragmatic appropriateness are influenced by rater variation, some studies instead conducted a seemingly more objective analysis of speech act realization strategies, predominantly using the coding system from the Cross-Cultural Speech Act Realization Project (CCSARP; Blum-Kulka, House, & Kasper, 1989). In native speaker data on requests (examined in our study), Blum-Kulka et al. identified nine head acts (i.e., the request proper) with three levels of directness: direct, conventionally indirect, and nonconventionally indirect (see Appendix A for our adaptation of their categories). Additionally, according to Blum-Kulka et al., requests can be mitigated by lexical and syntactic downgraders, such as "I was wondering if you could" and "Do you think you could," as well as supportive moves, like disarmers (e.g., "I know you are busy") and preparators (e.g., "Do you have a

minute?”). As Blum-Kulka et al. found, and following theoretical predictions (Brown & Levinson, 1978; Thomas, 1995), speech acts tend to be more face-threatening when there is social distance and a difference in power between interlocutors, and when the degree of imposition is high. To minimize the threat, speakers of various languages, including American English, prefer conventionally indirect strategies and mitigated devices.

Early SL studies using the CCSARP framework showed that in highly face-threatening requests, learners of lower proficiency, unlike native speakers, tend to use direct strategies due to their lack of linguistic resources, while those with higher proficiency use more conventionally indirect requests, mitigated expressions, and supportive moves, thus approximating NS patterns (e.g., Rose, 2000; Takahashi, 1996; Trosborg, 1995). More recent studies also observed developmental patterns between lower- and higher-level learners. In Otcu and Zeyrek (2008), advanced learners used more internal modifications (i.e., lexical and syntactic mitigated devices added to head acts) and external modifications (i.e., supportive moves) than lower-level learners (see also Félix-Brasdefer, 2007, and Göy, Zeyrek, & Otcu, 2012). However, as compared to NSs, advanced learners, while using external modifications at the same rate, underused internal syntactic modifications, suggesting that linguistic difficulties are at play. Other studies similarly show that advanced learners still do not approximate NS patterns in terms of frequency of strategy use (e.g., Achiba, 2003; Hendriks, 2008; Taguchi, 2011; Vilar-Beltran, 2008). For example, they may use more grounders than NSs (Faerch & Kasper, 1989) do, but fewer preparators (Taguchi, 2012).

Although pragmatic strategies are widely used to evaluate learners' speech act performance in comparison to that of NSs and to show patterns in their developmental progression, Taguchi (2006) questioned mainly using this framework and compared how well judges' ratings, as opposed to learners' inventories of pragmatic strategies, discriminate between different proficiency levels. She found that a judgment of overall appropriateness differentiated well between two proficiency levels, but there was only a marginal difference in pragmatic strategies used by the two groups. She suggested that additional aspects of performance that raters notice, such as grammatical and discourse control, also play a role in pragmatic appropriateness. Taguchi (2011, 2012) and Mori (2009) propose that conversation management skills and the ability to accomplish mutual understanding and resolve communication breakdowns can likewise affect pragmatic appropriateness; thus, assessing pragmatic performance only via strategies employed is overly simplistic.

NS Norms

The two methods of evaluating SL speech act performance, via pragmatic strategies or raters' judgments, normally involve comparisons to NS norms;

however, several researchers problematize this practice (Kasper & Schmidt, 1996; Roever, 2011; Taguchi, 2011). Kasper and Schmidt argue that

simply identifying differences [between NSs and NNSs] does not inform us which of those differences may matter in interaction. Some differences between NS norms and L2 performance may result in negative stereotyping by NS message recipients, whereas others may be heard as somewhat different but perfectly appropriate alternatives. (p. 56)

Hendriks (2010) responded to Kasper and Schmidt's call and experimentally examined how NSs evaluate e-mail requests made by NNSs with different amounts of syntactic modification. She found that NSs did not differentially evaluate requests with *can* vs. *could*, although according to Brown and Levinson (1987), requests modified with *could* are arguably less face-threatening. Hendriks explained that although, in judgment tasks asking NSs to order various isolated expressions according to appropriateness, *could you* requests were marked more appropriate than *can you* (e.g., Tanaka & Kawade, 1982) requests, in extended discourse in her study syntactic modifications such as *could you* alone might not have attracted NS attention. However, requests prefaced by *I was wondering if* were rated more positively than those with *can you*, suggesting that NSs may more readily notice highly involved modifications.

Hendriks' (2010) study shows that NSs may not orient to all modifications made by NNSs that are theoretically deemed to affect appropriateness (based on the examination of NS norms), which warrants the following question: Do learners need to use NS patterns in order to appear pragmatically appropriate, or can they develop unique styles that are nevertheless viewed as pragmatically appropriate by the target language speakers? After all, NSs are not uniform themselves. For example, Meier (1998) catalogued conflicting results of studies on NS apologies, indicating that preferred strategies differ based on social distance, gender, and specific situations. Ishihara (2013) similarly critiqued the use of NS norms and standardization of ratings in pragmatic assessment:

In authentic contexts, the learners' interactants are their real language appraisers, and they may not necessarily share a single yardstick. They are likely to assess learners' language use from a range of subjective perspectives, and they usually will not undergo rater training or norming. (p. 125)

Some NSs may even expect language learners to follow foreigner-specific rather than NS norms (Hassall, 2004; McNamara & Roever, 2006) and will overlook their inappropriate behaviour (Ishihara & Tarone, 2009). As Roever (2011) concluded, "the current reliance on narrow benchmarking [NS] samples deserves critical examination" (p. 13).

Research Questions

Both ratings of pragmatic appropriateness and examination of pragmatic strategies used by NSs versus NNSs present challenges for assessing pragmatics. Although the rating of pragmatics is even more subjective than that of other aspects of L2 performance (Ishihara, 2013), examination of pragmatic strategies alone misses other aspects that can affect pragmatic appropriateness (Taguchi, 2011). There appears to be an overreliance on NS norms in evaluating pragmatic appropriateness, and thus more studies are needed on what is salient to native speakers and raters when evaluating nonnative speaker pragmatic performance. We set out to examine raters' behaviour. To further examine the benefits and challenges of the two approaches (ratings versus pragmatic strategies), we compared rater judgments with the categorization of discourse by pragmatic strategies (following Taguchi, 2006), asking the following research questions:

1. How do raters orient to specific pragmatic strategies deemed to be more or less appropriate in prior research based on NS norms (e.g., Blum-Kulka et al., 1989; Taguchi, 2012)?
2. What factors are salient to raters when scoring pragmatic appropriateness of requests?
3. What are the sources of disagreement between raters?

Methodology

Participants

ESL students. The participants were 32 ESL learners in the United States enrolled in Level 2 courses (equivalent to an intermediate proficiency level) at a four-level pre-academic ESL program. The participants' average age was 20; there were 20 males and 12 females; the majority of students were NSs of Mandarin Chinese, with the remaining languages being Arabic, Korean, Cantonese, and Russian.

Raters. The authors of this article were the raters. The first rater was a female NNS of English, with four years of experience teaching ESL in the United States. The second rater was a male NS of English, with more than five years of experience teaching ESL in the United States. The third rater was a female NS of English with six years of experience teaching English in the United States, Asia, and Europe. All raters were in their 30s.

Ratings were completed as part of a larger study (Sydorenko, 2011); the raters did not familiarize themselves with, and thus were not influenced by, the literature on rater behaviour prior to the rating process. However, all raters were familiar with the literature on speech acts, speech act realization strategies, and conversation analysis (similar to the raters' background in Walters, 2007). Because we were interested in examining the behaviour of

raters who are well-versed in the literature on pragmatics, the inclusion of both native and near-native speaking raters was motivated by practical concerns of finding raters with such a background. We did not explicitly focus on comparing the judgments of native and near-native raters (due to the unbalanced number of them), but rather took a data-driven approach to examining disagreements between raters.

Instrument: Computerized Extended DCT with Video Prompts

To respond to Kasper's (2006), Roever's (2011), and Yates' (2010) call, this study examined extended discourse. However, although the above-mentioned researchers envisioned extended discourse as a product of real interlocutors' interaction, we examined "simulated" extended discourse, which resulted from participants' use of computer-delivered DCTs with multiple rejoinders that simulated a conversation. We did not opt for a natural conversation because it does not lend itself to comparisons across learners, although it is the most authentic of all data collection methods. Role-plays approximate authentic discourse because they allow for talk-in-interaction and discourse sequencing (Walters, 2007), but interlocutors can influence the progression of interaction and may affect participants' ratings (Roever, 2011). DCTs have been used to maximize standardization among learners, yet they produce the least authentic data (e.g., Golato, 2003). To deal with the weaknesses of classic DCTs and role-plays, Bardovi-Harlig and Shin (2014) discuss a new breed of DCT: a computer-delivered timed aural-oral DCT in which the prompt is presented aurally rather than in written form and learners respond orally; this arguably increases authenticity as learners do not plan their responses. We contend that in the present study we used a type of DCT that produces even more authentic discourse: a computerized oral DCT with three rejoinders (i.e., follow-up moves after the participants' responses) and video support. The rejoinders were not visible all at once but were rather presented one by one, as video recordings, after each response, which moves such "simulated extended discourse" closer to natural conversations and extended discourse than the presentation of all rejoinders at once.¹ The rejoinders allowed participants to engage in the negotiation of a request, while video recordings of initial prompts and rejoinders featuring a NS theoretically increased authenticity. In fact, as reported in Sydorenko (2011), several learners stated that these computerized extended DCTs appeared quite real to them.

The ratings of 64 samples were analyzed: 32 participants completed a DCT on a pretest and on a posttest (due to the design of the larger study, Sydorenko, 2011). The DCT scenario was designed to elicit a request from a student to an instructor for help with homework. The exact wording of the scenario was: "Ask your instructor for help on homework. The instructor is busy and even had to cancel office hours this week. However, this homework is very difficult for you and you cannot do it without help." We considered this to be a high-imposition, and thus a highly face-threatening, request due

to the unavailability of the instructor and the student's urgent need for help. The initial prompt and rejoinders for this DCT are provided in Appendix B.

Rating Scale

Like Alemi and Tajeddin (2013) and Walters (2007), we were interested in how raters approach the process of rating pragmatic appropriateness without specific training or guidelines. Thus, like Walters, we used a very general holistic rating scale without scale descriptors. Walters used a 4-point scale, but we hazarded that a scale with more points, as in Taguchi (2006), would be more useful in identifying any gains between the pretest and the posttest (the goal of Sydorenko, 2011). Like Taguchi, we defined pragmatic appropriateness as "the ability to perform speech acts appropriately according to situations" (p. 519). Our rating scale ranged from 1 (*extremely inappropriate*) to 6 (*completely appropriate*). Each rater used ratings ranging from 2 to 6.

Following Taguchi (2011), raters discussed the responses for which the ratings differed by more than one point and adjusted their ratings accordingly (five cases total). Initial interrater reliability was .95, and after discussing disagreements the interrater reliability was .97 (Cronbach's alpha using intraclass correlations, following Larson-Hall, 2010). This very high reliability, as compared to other studies, might be due to similar backgrounds of raters (all had knowledge of pragmatics, have taught ESL, and were of similar ages). In other studies, lower reliabilities were reported: .76 in Walters (2007), where specific training was not provided, and .92 in Taguchi (2011), where training was provided. However, an additional factor is the task being rated: in the larger study of Sydorenko (2011), interrater reliabilities as low as .81 were obtained for some tasks.

Comments by Raters

Following Walters (2007), raters provided comments on each response explaining their score. However, to collect more specific information on the rating process and decisions, raters also coded each element of the whole response (head acts, supportive moves, lexical and syntactic downgraders and upgraders, and any other salient features) as either appropriate or inappropriate.

Analysis

Two of the authors coded all request strategies and modifications in the data and resolved the discrepancies through discussion. After discussion, 100% agreement was reached. (Because the coding scheme evolved iteratively, with new categories added based on such discussions, percent agreement before these discussions could not be calculated.)

The Blum-Kulka et al. (1989) framework for coding requests was used, with some modifications informed by Taguchi (2012) and Schauer (2006).

Strategies not found in existing studies of requests, such as agreements and disagreements after the denial of a request, were added, based on iterative discussion between two raters (see Appendix A for a full list of strategies coded). There were two categories of strategies: head acts and supportive moves (the latter included moves accompanying the request proper and follow-up moves).

Raters' markings of appropriate and inappropriate strategies were compared to the Blum-Kulka et al. (1989) framework of direct, conventionally indirect, and nonconventionally indirect head acts, as well as mitigating and aggravating supportive moves. Following Blum-Kulka (1987), Brown and Levinson (1978), Thomas (1995), and Trosborg (1995), it was hypothesized that direct head acts would be marked inappropriate and conventionally indirect strategies would be marked appropriate, but hints could be perceived as either appropriate or inappropriate. We also expected that mitigating supportive moves (e.g., preparators, disarmers) catalogued in Blum-Kulka et al. (1989) would be marked as appropriate, while aggravating supportive moves (e.g., threats, moralizing) would be perceived as inappropriate. Strategies not mentioned in Blum-Kulka et al., such as an additional request or closing, were hypothesized to be either mitigating or aggravating, based on how they contribute to the imposition (see Table 2). We also conducted a content analysis of raters' comments to identify aspects of pragmatic performance salient to them. We discuss commonalities as well as individual differences.

Results and Discussion

The quantitative and qualitative results, as well as their discussion, are presented together according to the overall themes found.

Comparison of Hypothesized and Rater-Marked Appropriateness of Head Acts

We first examined instances of strategies marked as appropriate and inappropriate and compared our findings to theoretical claims of what makes requests more or less appropriate. According to Blum-Kulka (1987), NSs prefer conventionally indirect strategies in high-imposition requests, while direct strategies, and sometimes hints, often contribute to inappropriate requests. This trend also surfaced in our data, with several nuances worthy of discussion.

First, when examining the marking of imperatives (see Table 1), we were surprised that one of two instances was marked as appropriate even though this is considered to be the most direct and face-threatening strategy (Blum-Kulka et al., 1989). In the case marked as appropriate (provided in Example 1 below), the twice-used imposition minimizer "if you have a time" and the query preparatory "can you help me with the this homework" preceding the

imperative were coded as appropriate, probably masking the imperative. The inappropriate case “please help me to do the homework” did not include such supportive moves. The imperative is in bold.

- (1) *Um if you have a time uh can you help me with the this homework? It is very fit? difficult for me and I can't answer uh the question. If you have a time, uh please **help me to do this**.*

Next, want statements, all of which were unmitigated, were marked as inappropriate, except for one instance in which the raters disagreed. One rater marked “I want you to explain by your um explain it to me” as appropriate, possibly because it was surrounded by other strategies that were perceived as appropriate (similar to Example 1): agreement with the instructor’s suggestion preceded the want statement, and the imposition minimizer “I can choose another time to come to your office” followed the want statement.

Table 1
Marking of Head Acts by Raters

	<i>Appropriate^a</i>	<i>Inappropriate^a</i>	<i>Disagreement</i>
<i>Direct</i>			
Imperative	1	1	
Want statement		8	1
Need statement	13	4	2
Statement of future action		1	
Hedged performative	2		
<i>Conventionally Indirect</i>			
Query preparatory			
can	10	2	3
could	7		
would			1
Permission	4		
Mitigated expression	11	1	2
<i>Nonconventionally Indirect</i>			
Hint	19	6	1

Note. The table represents counts.

^aCounted when at least one rater marked a strategy as appropriate or inappropriate.

Thus, while our data on imperatives and want statements conform to Blum-Kulka’s (1987) findings that direct head acts tend to be inappropriate in high-imposition requests with power difference, extended discourse surrounding these strategies can also play a role.

Need statements also appear to fall under Blum-Kulka et al.’s (1989) category of want statements (see also Taguchi, 2012), but we coded need state-

ments separately as a direct type of strategy. Need statements were marked more often as appropriate than as inappropriate (14 versus 4 instances respectively), in contrast to want statements. It might thus be worthwhile in future studies to examine want and need statements separately, as they seem to display different levels of directness.

We also had one instance of a statement of future action: "I will to go there ask you something about that." This strategy does not appear in Blum-Kulka et al. (1989), probably because NSs do not use it in requests. We coded it as a direct strategy because it was a statement rather than a question. As predicted, this direct strategy was marked as inappropriate.

Finally, there were two instances of hedged performatives. Even though, according to Blum-Kulka et al. (1989), this is a direct strategy both instances were marked as appropriate. It is premature to draw any conclusions based on only two instances, but some possibilities are that, as with need statements, reconceptualization of the directness level of this strategy might be necessary, or the context in which hedged performatives appeared in our study made them sound appropriate.

Next, we examined the raters' markings of conventionally indirect strategies. Query preparatories with *could* were marked as appropriate in all cases; those with *can* were marked as appropriate in 10 cases, inappropriate in 2; and there was disagreement in three instances. A query preparatory with *would* only appeared once, and raters disagreed on its appropriateness. Our results support those of Hendriks (2010), who found that NSs did not differentially rate e-mails with *can you* and *could you*; in our study, many instances of *can* were rated as appropriate. Our explanation is the same as Hendriks': in extended discourse, a variety of other aspects of speech act performance can be more salient than individual expressions.

As expected, questions about permission were marked as appropriate. For mitigated expressions, however, 1 out of 14 was marked as inappropriate, and there was disagreement on 2 of them. We found this surprising because NSs overwhelmingly use mitigated expressions in appropriate responses (e.g., Taguchi, 2012). The mitigated expression "I wish you can help me to finish my homework" was marked inappropriate: it is not a target-like expression and probably was not interpreted as a mitigated expression by the raters. The raters disagreed on both occurrences of the expression "Could you do me a favor to help me," which the NNS rated as inappropriate and the male NS rated as appropriate.

The fact that six hints were marked as inappropriate also supports Blum-Kulka (1987) in that hints, although considered the most indirect strategies, are not always appropriate due to extreme vagueness.

Thus far, our findings indicate that raters are in agreement with other studies on the connection between pragmatic appropriateness and types of head acts used, but with some individual variation and influences of surrounding discourse.

Comparison of Hypothesized and Rater-Marked Appropriateness of Supportive Moves

When examining supportive moves (see Table 2), the same pattern emerges: while theoretically mitigating and aggravating moves are generally marked as such, this is not always the case. For example, 10 out of 33 grounders, 6 out of 24 imposition minimizers, and 2 out of 10 disarmers were marked as inappropriate. Sometimes these strategies were in fact used inappropriately, such as the grounder “This weekend I have extremely busy schedule,” on which one rater commented “Get in line—the professor is busy too, that’s why they cancelled the office hours.” However, most of the time these strategies on their own seemed appropriate, but they appeared in the context of inappropriate statements. In Example 2 below, the minimizer “Uh I think I just need a little a little time um to for your help” does not sound inappropriate; however, it was marked as such, possibly because it was preceded by the

Table 2
Marking of Supportive Moves by Raters

	<i>Appropriate</i>	<i>Inappropriate</i>	<i>Disagreement</i>
<i>Potentially Mitigating</i>			
Preparator	20	1	2
Grounder	33	10	1
Imposition minimizer	24	6	1
Appealer	5	2	1
Disarmer	10	2	
Promise	4		
Confirmation	4	1	
Appreciation	133	1	
Apology	11		
Agreement	73	11	2
Offer of solution: no hearer effort required	5		
Suggestion	4	2	2
Concession	10		
Sweetener	3	1	
Additional request: for information	8	1	
Resolution	13	7	
Farewell/closing	3	2	1
<i>Potentially Aggravating</i>			
Moralizing		2	
Threat		2	
Repeat request	5	6	2
Disagreement	9	28	1
Additional request: for time	5	17	3

Note. The table represents counts.

need statement, the grounder, and the imperative all marked as inappropriate. Again, there is evidence that raters were looking at the appropriateness of each expression in context rather than in isolation.

- (2) *Uh but I really need your help. Uh if you um can't help me I really uh don't know how to do it. Um so please help me to do the homework. Uh I think I just need a little a little time um to for your help.*

Interestingly, there was even one inappropriate appreciation statement. One learner used the phrase "Thank you all the same" on the pretest and on the posttest. The pretest instance was marked as inappropriate, with one rater commenting, "Something about that phrase just rubs me the wrong way. It's like a sullen 'Well, that was useless, but thanks anyway.'" However, on the posttest, it was marked as appropriate. The context surrounding the utterance is important to consider. In both cases, the utterance is preceded by an indication of effort to resolve the problem without the instructor's help, and if that fails, e-mailing him (as the instructor suggested). However, there are differences in what appears right before the "Thank you" utterance. In the inappropriate case, it is a pause followed by *but*, which aligns with prior findings that pauses and markers of disagreement, like *but*, can accompany dispreferred responses (Schegloff, 2007). In the appropriate case, there is no pause, and *and* rather than *but* is used. This example is yet another support for our finding that discourse plays a large role in perceptions of appropriateness, and that examining only the types of strategies learners use is insufficient.

To summarize our findings on assessing pragmatics via strategies used, we question if categorizing the strategies as mitigating and aggravating, as Blum-Kulka et al. (1989) did, is appropriate for examining NNS data. It may be the case that grounders and imposition minimizers are always mitigating when used by NSs, but due to NNSs' linguistic difficulties and cultural differences, strategies intended to be mitigating may not always be perceived as such. As will be shown below, this caused the most disagreement between raters because they had to interpret how an inappropriate-sounding strategy was actually intended to function. We thus contend that the commonly practiced comparison of frequencies with which NSs and NNSs use various strategies can be problematic because important details, such as whether mitigating strategies were actually used appropriately by the learners, can be missed.

Rater Orientation to the Percentage of Appropriate Strategies

We were also intrigued by the fact that context mattered more than individual strategies to the raters, and decided to examine this in a different way. We wanted to see if mitigated expressions, such as "I was wondering if you could," correlate with high ratings, our reasoning being that learners rarely use mitigated expressions, while NSs overwhelmingly do (Taguchi, 2012). We thus hypothesized that only high-scoring learners would use these ex-

pressions and thereby approximate NS patterns. Such a relationship was not found: learners who used mitigated expressions achieved ratings between 3 (somewhat inappropriate) and 6 (very appropriate). Example 3 below shows one response that includes a mitigated expression and received ratings of 3 and 4. The expression marked as inappropriate (i.e., the moralizer) is underlined, and the mitigated expression is in bold. One rater's comment was "Intonation very hurried, request is somewhat frustrated/demanding." This learner made other inappropriate responses in the follow-up moves, such as "I hope you can response me, OK?" While marking the mitigated expression as appropriate, the raters did not make any specific comments about it, which indicates that it was not very salient in light of the rest of the response.

- (3) *Hey uh the reason I come here is that I want to talk about homework. Uh you know this weekend I have extremely busy schedule. And I found I uh last weekend I I I already come here to find you talk about this but you cancel your office hours this weekend so I afraid I I have to talk this because I think uh homework is very difficult for me and I can't do it without help. So uh I was won I was wondering do you have free time?*

In fact, 14 out of 19 learners who achieved a high rating (above 5) did not use any mitigated expressions, which suggests to us that too much emphasis has been placed in the literature on the importance of teaching mitigated expressions (see Taguchi, 2012, and Woodfield, 2013, for a review). Although mitigated expressions are overwhelmingly used by NSs, it appears that raters do not think learners need to use them to appear pragmatically appropriate. Instead, we found that judges oriented to the ratio of appropriate and inappropriate strategies in a given response. For example, one participant, who used two want statements that were rated inappropriate, achieved an average rating of 5.7. One rater wrote, "Only the beginning is a little 'bumpy,' but the rest is excellent." This comment indicates that raters oriented toward the whole request sequence, including the closing of the interaction, rather than specifically on the request proper, and that, in the raters' minds, some inappropriate strategies could be compensated for by appropriate ones if there were enough of the latter type. However, the SL pragmatics literature overwhelmingly focuses on narrow aspects of speech acts rather than on the more complete negotiation sequences. (Evidence of this can be found in the fact that we had to develop codes for various follow-up moves not discussed in prior research.)

Rater Orientation to Intonation

We also noticed that raters sometimes coded overall intonation or intonation on specific expressions as appropriate (six instances) or inappropriate (also six instances). For example, one comment was "Would be more polite with peppier intonation," which again indicates that the use of specific strategies and expressions cannot on its own determine the appropriateness of

a response. Although Blum-Kulka et al. (1989) include intonation in their “orthographic/suprasegmental emphasis” category, it is only conceptualized as an aggravating supportive move. However, intonation can be mitigating (Félix-Brasdefer, 2012; Flores-Ferran, 2012), as was also found in our study. Examination of nonverbal features is rare in SL pragmatics studies (but see Gass & Houck, 1999), and we concur with Félix-Brasdefer and Koike (2012) that more emphasis should be placed on examining the connection between pragmatics and suprasegmental features.

Rater Orientation to Repetitiveness

Another observation was that raters viewed repetitiveness negatively. For example, one learner used the cajoler *you know* seven times. The first three cajolers were marked appropriate, while the next four were marked inappropriate. One rater commented, “What makes these inappropriate is the number.” Repetition of the original request was also generally inappropriate, while other requests, suggestions, or solutions tended to be viewed more favourably. Many studies have found that language learners, as opposed to NSs, display a smaller repertoire of strategies (Félix-Brasdefer, 2007; Göy et al., 2012), but what effect this has on raters has not been examined. Based on our findings, we stress that variability in strategies has a significant impact on the perceptions of pragmatic appropriateness.

Rater Orientation to Cultural Misunderstanding

Raters also expected the learners to understand cultural concepts; when misunderstandings occurred, they were marked as inappropriate. The most common misunderstanding in our data concerned the role of a tutor assigned to a course. Many participants questioned the tutor’s competence, some were concerned that they would have to pay the tutor, and some asked for contact information. This was always manifested as a disagreement with an instructor’s suggestion, which was marked as inappropriate. One learner said, “Do you think the tutors can help me? I mean I’m not [...] to trust the tutor.” This finding supports Meier’s (2010) claim that cultural knowledge is vital in achieving pragmatic appropriateness and should be accounted for in pragmatics instruction. In our case, most learners probably would not have disagreed with an instructor’s suggestion had they understood the concept of a tutor assigned to a course.

Disagreements Between Raters

We now turn to examining disagreements between raters. Although disagreements were few (9% for head acts, 3% for supportive moves), it is nevertheless worth examining what, on rare occasions, caused raters to disagree. The two most common disagreements were based on (a) individual raters’ ideas of what would be an appropriate strategy in a specific situation, and (b) differences in interpreting the intended meanings of non-target-like expressions.

Regarding the first pattern, one example is raters' disagreement on the appropriateness of the content of a student's suggestion to an instructor. To get help with homework, one participant said, "Maybe we can uh uh we can eat eat eat lunch together. Then I can ask you a little more ti a little time." The NS male rater coded this as appropriate and made the comment "Clever!" However, the other two raters marked this as inappropriate. Sometimes raters also disagreed on how much face work is needed before an actual request. The following response was marked as appropriate by two raters, and inappropriate by the female NS: "Oh you know uh I've got some problem with the homework. Uh can I get some help?" The female NS rater commented, "I have the impression that the speaker is trying to be appropriate, but is still a little gruff, a little short." These examples indicate that individuals have their own interpretations of appropriate and inappropriate behaviour, and there is no one target-language norm for every aspect in pragmatics. Knowing one's interlocutor and their preferences well is also an important consideration for achieving pragmatic appropriateness with a given person.

Awkwardness of expressions also seemed to interfere with judgments of pragmatic appropriateness, often causing disagreements between raters of what the intended message was. The closing statement "And I am waiting for your answer" was rated as inappropriate by the male rater, who wrote, "From his comment, it sounds like he still believes that the responsibility for helping him is solely his teacher's, and he has no intention of trying to find the tutor." However, both female raters thought this was appropriate, although linguistically awkward. One of them wrote, "This seems like a misapplied strategy, like something this person learned to put in a letter or e-mail and is using here but without a solid reason for doing so." Similarly, the raters disagreed on the person's intentions regarding the statement "Maybe it it will cost cost you many times to help me." Two raters thought this was a linguistically awkward way of minimizing imposition, while the NS female rater wrote, "This is the opposite of an effort to minimize imposition, as the speaker openly acknowledges the imposition." That is, one rater decided to examine the literal meaning, while the other two thought it would be illogical for this person to increase the imposition and therefore this was not the student's intent. It appears that raters viewed the participants differently: some treated them like linguistically deficient NNSs and thus gave them the benefit of the doubt, leaning toward considering inappropriate statements as linguistically awkward instead. Others tended to view statements appearing to be inappropriate as intended to be such, rather than caused by learners' linguistic deficiencies. However, we found that this was not a pattern for specific raters, but varied for each expression that raters evaluated.

Conclusion

One of our findings is that raters who are well versed in pragmatics literature tend to view strategies theorized to be typically appropriate or inappropriate as such. However, context plays a vital role, as the exact same expression can be marked as either appropriate or inappropriate by the same rater, depending on the context. Although it is common knowledge that not only the directness of the head act, but also internal and external modifications of it, influence the overall appropriateness of the speech act, literature on SL pragmatics (teaching, research, assessment) tends to examine these separately rather than contextually. We showed the problems that one can encounter in taking such an approach.

Additionally, our findings suggest that there is an overemphasis in SL pragmatic literature on pragmatic strategies and expressions. On the other hand, we identified additional aspects that raters orient to when judging pragmatic appropriateness—ratio of appropriate to inappropriate strategies used, repetitiveness, intonation, cultural knowledge—and like Mori (2009), Roever (2011), and Taguchi (2012), we argue that pragmatic competence should be conceptualized more broadly.

Finally, it is also vital to know and remember that learners do not need to behave exactly like NSs to be pragmatically appropriate. Some raters (but not all) will “give them the benefit of the doubt” and assume that some awkward expressions, although sounding inappropriate, were intended as appropriate. Moreover, rater agreement was high but there were some disagreements, indicating that although NS norms do not exist, there are tendencies.

As with most studies, ours has limitations. To allow for an intensive analysis, we examined one situation (a request to an instructor), and thus some of our specific findings, such as the effect of repeated requests, might not extend to other situations. However, other findings, such as the importance of understanding culture, variability of pragmatic strategies and linguistic expressions, and using certain expected intonation, can be generalized to other situations. Another limitation is that we do not have specific suggestions for what makes intonation appear appropriate or inappropriate, as we did not analyze intonation contours. Raters instead commented on intonation impressionistically. Future studies should examine whether there are specific intonation contours or other prosodic features that appear appropriate or inappropriate. One additional limitation is the specific population of raters: all had experience teaching ESL and/or EFL, all were in their 30s, and all were familiar with literature on pragmatics; thus they likely approached the ratings differently from NSs with other characteristics (different professions, generations, experience, etc.). Future studies should investigate how various native and near-native speakers, who know or do not know the learners, orient to rating their responses in terms of pragmatic appropriateness, and what aspects they consider most salient.

Classroom Implications

Our first teaching suggestion is that there should be a balance of general and specific. That is, learners should be exposed to patterns (or tendencies) common for NSs (following Yates, 2010), but they should also be shown counter-examples and contexts in which those tendencies do not work. For example, learners can be presented with behaviours typically considered appropriate by instructors in academic situations, such as giving enough advance notice, but they should also know that this varies depending on the instructors' characteristics, their relationship with particular students, their mood that day, and so on. Learners should be encouraged to look for opportunities to interact with various native speakers in order to get a feel for the existence of patterns, as well as a great deal of variability. In the same vein, learners should be presented with a bird's-eye view. As Meier (2010) states, given that there are too many peculiarities specific to each situation, it may be more fruitful to help learners understand more broadly the cultural canvas that mediates what is pragmatically appropriate in each situation. For example, giving advance notice can be presented as an overall cultural value, not something that is specific to instructors' expectations.

Next, given that there are so many nuances to pragmatic competence that may take a long time to develop in an L2, we suggest that, as more research becomes available, teachers focus on the most salient aspects affecting target language speakers' judgments. We by no means discount the current emphasis on drawing learners' attention to narrow aspects of pragmatics, such as mitigated expressions, but rather suggest that other aspects of pragmatic competence, such as variability of strategies and expressions and nonverbal communication, should be also part of pragmatic instruction. It may also be helpful to present learners with research findings indicating what target speakers consider to be "grave mistakes" of pragmatic appropriateness, versus which ones are typically forgiven, so that students can prioritize their learning. Our study suggests that such patterns may exist.

Implications for Testing

The main implication for testing from our study is that reliance on NS norms does need to be problematized, as other researchers, such as Ishihara (2013) and Roever (2011), also contend. Although, as our study shows, raters may show a high degree of agreement, possibly due to their similar backgrounds, there can also be some unresolved disagreements. Taguchi (2011) similarly observed that raters do not always agree with the criteria provided to them during training and may continue to rely on their own. Using an example from our study, to maximize consistency all raters could be given a "key" of correct responses (e.g., the fact that suggesting a lunch meeting with an instructor should be marked as inappropriate), based on the majority of re-

sponses of NS instructors polled earlier. However, how meaningful is such standardization when a standard does not truly exist? And, ultimately, how useful is this type of approach for language learners who need to know that variation exists among NSs or near-native users of the target language? Thus, how raters are trained, how agreement is reached (or not), and how correct answers are established should be examined more critically.

Acknowledgments

This study was funded by (a) the U.S. Department of Education International Research and Studies (IRS) Program Grant, special project number P017A100100, and (b) Dissertation Completion Fellowship from Michigan State University. We would like to thank the technical team led by Dr. Hoopingarner and the research participants for help with various aspects of this project.

The Authors

Tetyana Sydorenko is an Assistant Professor of Applied Linguistics at Portland State University. Her research interests include teaching and acquisition of SL pragmatics, computer-assisted language learning, psycholinguistic processes in SL acquisition, and assessment. She is currently investigating the use of adaptive computer-simulated conversations in the teaching of SL pragmatics.

Carson Maynard (English Language Institute, University of Michigan) has a BA in Linguistics from Duke University and an MA in Linguistics from the University of Michigan. His professional interests include pronunciation, conversation analysis, teacher training, and integrating technology into language learning.

Erin Guntly completed a BA in Applied Linguistics at Portland State University and is currently a PhD student at the University of British Columbia. Her research focuses primarily on the syntax and semantics of discourse, particularly in understudied languages of North America and Africa.

Note

1 As described in Sydorenko (2011, in press), extensive piloting of these computerized DCTs was conducted, which uncovered the most common progressions of talk in the given situations. The prompts and their sequencing were based on this extensive piloting; this process helped minimize the mismatches between participants' responses and following prompts. Surprisingly, in the present study for the DCT examined, no mismatches occurred.

References

- Achiba, M. (2003). *Learning to request in a second language: A study of child interlanguage pragmatics*. Clevedon, UK: Multilingual Matters.
- Alemi, M., & Tajeddin, Z. (2013). Pragmatic rating of L2 refusal: Criteria of native and non-native English teachers. *TESL Canada Journal*, 30(7), 63–81.
- Bardovi-Harlig, K., & Shin, S.-Y. (2014). Expanding traditional testing measures with tasks from L2 pragmatics research. *Iranian Journal of Language Testing*, 4(1), 26–49.
- Blum-Kulka, S. (1987). Indirectness and politeness in requests: Same or different? *Journal of Pragmatics*, 11(2), 131–146. doi:10.1016/0378-2166(87)90192-5
- Blum-Kulka, S., House, J., & Kasper, G. (Eds.). (1989). *Cross-cultural pragmatics: Requests and apologies*. Norwood, NJ: Ablex.
- Brown, J. D., & Ahn, R. C. (2011). Variables that affect the dependability of L2 pragmatic tests. *Journal of Pragmatics*, 43(1), 198–217. doi:10.1016/j.pragma.2010.07.026
- Brown, J. D., Hudson, T. D., Norris, J. M., & Bonk, W. (2002). *Investigating task-based second language performance assessment*. Honolulu, HI: University of Hawai'i Press.

- Brown, P., & Levinson, S. C. (1978). Universals in language usage: Politeness phenomena. In E. N. Goody (Ed.), *Questions and politeness: Strategies in social interaction* (pp. 56–311). Cambridge, UK: Cambridge University Press.
- Faerch, C., & Kasper, G. (1989). Internal and external modification in interlanguage request realization. In S. Blum-Kulka, J. House, & G. Kasper (Eds.), *Cross-cultural pragmatics: Requests and apologies* (pp. 221–247). Norwood, NJ: Ablex.
- Félix-Brasdefer, J. C. (2007). Pragmatic development in the Spanish as a FL classroom: A cross-sectional study of learner requests. *Intercultural Pragmatics*, 4(2), 253–286. doi:10.1515/IP.2007.013
- Félix-Brasdefer, J. C. (2012). Pragmatic variation by gender in market service encounters in Mexico. In J. C. Félix-Brasdefer & D. Koike (Eds.), *Pragmatic variation in first and second language contexts: Methodological issues* (pp. 17–48). Amsterdam, Netherlands: John Benjamins.
- Félix-Brasdefer, J. C., & Koike, D. A. (Eds.). (2012). *Pragmatic variation in first and second language contexts: Methodological issues*. Amsterdam, Netherlands: John Benjamins.
- Flores-Ferrán, N. (2012). Pragmatic variation in therapeutic discourse: An examination of mitigating devices employed by Dominican female clients and a Cuban American therapist. In J. C. Félix Brasdefer & D. A. Koike (Eds.), *Pragmatic variation in first and second language contexts: Methodological issues* (pp. 81–112). Amsterdam, Netherlands: John Benjamins.
- Gass, S. M., & Houck, N. (1999). *Interlanguage refusals: A cross-cultural study of Japanese-English*. Berlin, Germany: Walter de Gruyter.
- Golato, A. (2003). Studying compliment responses: A comparison of DCTs and recordings of naturally occurring talk. *Applied Linguistics*, 24(1), 90–121. doi:10.1093/applin/24.1.90
- Göy, E., Zeyrek, D., & Otcu, B. (2012). Developmental patterns in internal modification of requests: A quantitative study on Turkish learners of English. In M. Economidou-Kogetsidis & H. Woodfield (Eds.), *Interlanguage request modification* (pp. 51–86). Amsterdam, Netherlands: John Benjamins.
- Hassall, T. (2004). Through a glass, darkly: When learner pragmatics is misconstrued. *Journal of Pragmatics*, 36(5), 997–1002. doi:10.1016/j.pragma.2003.09.001
- Hendriks, B. (2008). Dutch English requests: A study of request performance by Dutch learners of English. In M. Putz & J. N. Aertselaer (Eds.), *Developing contrastive pragmatics: Interlanguage and cross-cultural perspectives* (pp. 335–354). Berlin, Germany: Walter de Gruyter.
- Hendriks, B. (2010). An experimental study of native speaker perceptions of non-native request modification in e-mails in English. *Intercultural Pragmatics*, 7(2), 221–255. doi:10.1515/iprg.2010.011
- Ishihara, N. (2013). Teacher-based assessment of L2 Japanese pragmatics: Classroom applications. In S. Ross & G. Kasper (Eds.), *Assessing second language pragmatics* (pp. 124–148). Basingstoke, UK: Palgrave Macmillan.
- Ishihara, N., & Cohen, A. D. (2010). *Teaching and learning pragmatics: Where language and culture meet*. Harlow, UK: Pearson.
- Ishihara, N., & Tarone, E. (2009). Subjectivity and pragmatic choice in L2 Japanese: Emulating and resisting pragmatic norms. In N. Taguchi (Ed.), *Pragmatic competence* (pp. 101–128). Berlin, Germany: Walter de Gruyter.
- Kasper, G. (2006). Speech acts in interaction: Towards discursive pragmatics. In K. Bardovi-Harlig, C. Félix-Brasdefer, & A. S. Omar (Eds.), *Pragmatics and language learning* (Vol. 11, pp. 281–314). Honolulu, HI: University of Hawai'i Press.
- Kasper, G., & Schmidt, R. (1996). Developmental issues in interlanguage pragmatics. *Studies in Second Language Acquisition*, 18(2), 149–169. doi:10.1017/S0272263100014868
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York, NY: Routledge.
- Liu, J. (2007). Comparing native and nonnative speakers' scoring in an interlanguage pragmatics test. *Modern Foreign Languages*, 30(4), 395–404.
- Liu, J., & Xie, L. (2014). Examining rater effects in a WDCT pragmatics test. *Iranian Journal of Language Testing*, 4(1), 50–65.

- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Oxford, UK: Wiley-Blackwell.
- Meier, A. J. (1998). Apologies: What do we know? *International Journal of Applied Linguistics*, 8(2), 215–231. doi:10.1111/j.1473-4192.1998.tb00130.x
- Meier, A. J. (2010). Culture and its effect on speech act performance. In A. Martínez-Flor & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical and methodological issues* (pp. 75–90). Amsterdam, Netherlands: John Benjamins.
- Mori, J. (2009). Commentary: The social turn in second language acquisition and Japanese pragmatics research: Reflection on ideologies, methodologies and instructional implications. In N. Taguchi (Ed.), *Pragmatic competence* (pp. 335–358). Berlin, Germany: Walter de Gruyter.
- Otcu, B., & Zeyrek, D. (2008). Development of requests: A study on Turkish learners of English. In M. Putz & J. N. Aertselaer (Eds.), *Developing contrastive pragmatics: Interlanguage and cross-cultural perspectives* (pp. 265–299). Berlin, Germany: Walter de Gruyter.
- Roever, C. (2008). Rater, item, and candidate effects in discourse completion tests: A FACETS approach. In E. Alcón Soler & A. Martínez-Flor (Eds.), *Investigating pragmatics in foreign language learning, teaching, and testing* (pp. 249–266). Clevedon, UK: Multilingual Matters.
- Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing*, 28(4), 463–481. doi:10.1177/0265532210394633
- Roever, C., Fraser, C., & Elder, C. (2014). *Testing ESL sociopragmatics: Development and validation of a web-based test battery*. New York, NY: Peter Lang.
- Rose, K. R. (2000). An exploratory cross-sectional study of interlanguage pragmatic development. *Studies in Second Language Acquisition*, 22(1), 27–67. doi:10.1017/S0272263100001029
- Schauer, G. A. (2006). The development of ESL learners' pragmatic competence: A longitudinal investigation of awareness and production. In K. Bardovi-Harlig, C. Félix-Brasdefer, & A. S. Omar (Eds.), *Pragmatics and language learning* (pp. 135–163). Honolulu, HI: University of Hawai'i Press.
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis* (Vol. 1). Cambridge, UK: Cambridge University Press.
- Sydorenko, T. (2011). Exploring the potential of rehearsal via automatized structured tasks versus face-to-face pair work to facilitate pragmatic and oral development. (Doctoral dissertation, Michigan State University). *Dissertation Abstracts International*, 3464977.
- Sydorenko, T. (in press). The use of computer-delivered structured tasks in pragmatic instruction: An exploratory study. *Intercultural Pragmatics*.
- Taguchi, N. (2006). Analysis of appropriateness in a speech act of request in L2 English. *Pragmatics*, 16(4), 513–533. doi:10.1075/prag.16.4.05tag
- Taguchi, N. (2011). Rater variation in the assessment of speech acts. *Pragmatics*, 21(3), 453–471. doi:10.1075/prag.21.3.08tag
- Taguchi, N. (2012). *Context, individual differences and pragmatic competence*. Toronto, ON: Multilingual Matters.
- Tajeddin, Z., & Alemi, M. (2014). Pragmatic rater training: Does it affect non-native L2 teachers' rating accuracy and bias? *Iranian Journal of Language Testing*, 4(1), 66–83.
- Takahashi, S. (1996). Pragmatic transferability. *Studies in Second Language Acquisition*, 18(2), 189–223. doi:10.1017/S0272263100014881
- Tanaka, S., & Kawade, S. (1982). Politeness strategies and second language acquisition. *Studies in Second Language Acquisition*, 5(1), 18–33. doi:10.1017/S0272263100004575
- Thomas, J. A. (1995). *Meaning in interaction: An introduction to pragmatics*. London, UK: Longman.
- Trosborg, A. (1995). *Interlanguage pragmatics: Requests, complaints, and apologies*. New York, NY: Walter de Gruyter.
- Vilar-Beltrán, E. (2008). The use of mitigation in role-play activities: A comparison between native and nonnative speakers of English. In E. Alcón-Soler (Ed.), *Learning how to request in an instructed language learning context* (pp. 127–142). Berlin, Germany: Peter Lang.
- Walters, F. S. (2007). A conversation-analytic hermeneutic rating protocol to assess L2 oral pragmatic competence. *Language Testing*, 24(2), 155–183. doi:10.1177/0265532207076362

Walters, F. S. (2009). A conversation analysis-informed test of L2 aural pragmatic comprehension. *TESOL Quarterly*, 43(1), 29–54. doi:10.1002/j.1545-7249.2009.tb00226.x

Wolfe, E. W., & Chiu, C. W. T. (1997). *Detecting rater effects with a multi-faceted rating scale model*. Retrieved from <http://files.eric.ed.gov/fulltext/ED408324.pdf>

Woodfield, H. (2012). “I think maybe I want to lend the notes from you”: Development of request modification in graduate learners. In M. Economidou-Kogetisidis & H. Woodfield (Eds.), *Interlanguage request modification* (pp. 9–50). Amsterdam, Netherlands: John Benjamins.

Yates, L. (2010). Speech act performance in workplace settings. In A. Martínez-Flor & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical and methodological issues* (pp. 109–126). Amsterdam, Netherlands: John Benjamins.

Youn, S. J. (2007). Rater bias in assessing the pragmatics of KFL learners using FACETS analysis. *Second Language Studies*, 26(1), 85–163.

Appendix A

Categories for Coding Request Strategies

Strategy	Examples
	<i>Head Acts</i>
Imperative/mood derivable	Please help me to do the homework.
Want statement	I want you to help me with my homework.
Need statement ^a	I need your help with my homework.
Statement of future action ^a	I will to go there ask you something about that.
Hedged performative	I want to ask you to help me for my homework.
Query preparatory	Can you give me some helps?
Permission ^b	Can I go to your office during your office hour?
Mitigated expression ^b	I was wondering if you can help me.
Hint ^b	I can't do it without help.
	<i>Supportive Moves^c</i>
Preparator	I have some question about my homework.
Grounder	This homework is very difficult for me.
Imposition minimizer	If you have some free time to help me.
Appealer	And I'm depressed these days. (Unlike in Blum-Kulka et al., 1989, the utterances that qualify go beyond tag questions.)
Disarmer	I know you are very busy.
Promise ^b	It will be OK.
Confirmation ^b	You mean that, right?
Moralizing	I can't find you.
Threat	If you cannot help me, I cannot finish it.
Appreciation ^b	Thank you very much.
Apology ^b	I'm sorry to bother you.
Agreement ^a	Oh this is good idea.
Offer of Solution (no hearer effort needed) ^a	Then I will search the answer online and ask my friends to do that.

Concession ^a	I know you are busy. (Unlike disarmers, concessions are made after the request.)
Sweetener ^d	You are so nice.
Additional request: for information ^a	Can I just ask your cellphone number?
Additional request: for time ^a	And can I only visit you 15 minutes?
Resolution ^a	So I will use the PowerPoint you showed us.
Suggestion ^a	Maybe we can have another time to talk about homework.
Farewell/Closing ^a	OK, see you next week.
Repeat request	(The speaker repeats the initial request literally or by paraphrase.)
Disagreement ^a	But the tutor have to pay money.

Note. Unless otherwise stated, the strategies are based on Blum-Kulka et al. (1989).

^aDeveloped by the authors. ^bAdopted from Taguchi (2012). ^cMoves accompanying the initial request, as well as all follow-up moves. ^dAdopted from Schauer (2006).

Appendix B

Extended DCT Initial Prompt and Rejoinders^a

1. OK, what's your question?
2. Oh, I'm I'm really sorry, th this week is very bad for me. I'm extremely busy. And in fact that's why I had to cancel my office hours. What I'd recommend is that you talk to the tutor. He is very good at explaining these things.
3. I'm afraid I'm really pressed for time right now and in fact I've gotta go teach in a few minutes. And and I truly apologize but this week as a said is really busy. I'm I'm leaving town, so I'm not gonna be able to meet with you in person. But if you'd like to e-mail me, I'd be more than happy to uh to respond to you. And also you might wanna check the PowerPoints that we used in class. You could also use the software in the language lab.
4. Yeah, and we can talk during my office hours next week.

^aThe initial prompt and rejoinders were followed by participants' responses.